

**A PRACTICAL GUIDE TO HEAVY TAILS:
STATISTICAL TECHNIQUES
FOR ANALYZING HEAVY TAILED DISTRIBUTIONS**

**Robert J. Adler
Raisa E. Feldman
Murad S. Taqqu**

Editors

DTIC QUALITY INSPECTED 13

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

19980206 077

Preface

Ever since data have been collected, they have fallen into two quite distinct groups: "good data", which meant that their owner knew how to perform the analysis, and "bad data", which were difficult, if not impossible, to handle.

Since the development of modern statistics over half a century ago, good data were typically those whose distribution was amenable to the tools of the theory, tools which invariably assumed the distributions of a first course in statistics; i.e. normal, chi-squared, etc.

Whereas bad data came in many forms, one type tended to jump around too much and involved outliers which contained important information. This, in short, was data with heavy tailed histograms. Economists, for example, have been well aware for almost thirty years that much economic data falls into this "bad" category, as are modern financial analysts. Data of this kind, however, arise in a far wider variety of fields than the economic, including statistical physics, automatic signal detection, and telecommunications, to name just three.

What made this type of data "bad", however, was nothing intrinsic, but rather the absence of well developed statistical techniques for its analysis.

Heavy-tailed distributions and processes have been studied for decades by probabilists and mathematical statisticians, with the last decade or so having seen major advances. Many of these are summarised in the 1994 monograph on *Stable Non-Gaussian Random Processes* by Samorodnitsky and Taqqu, which provides a theoretical background to the papers in this volume. The current collection, however, is directed to the general practitioner and is primarily concerned with techniques for data analysis.

Interestingly, despite the lack of a large-scale coordinated effort to develop techniques for the analysis of heavy-tailed data, it turns out that there are really a good number of them, scattered through a variety of different disciplines. It was in an attempt to bring together these various disciplines, and to "compare notes", that a small workshop was held in Santa Barbara in December 1995, with ONR support, and it was from the success of that workshop that the current volume grew.

We set about collecting expository papers on applications, data analytic techniques, and models, for heavy-tailed distributions and processes. We, and our authors, worked hard to write in a style easily accessible to readers in different disciplines. In fact, our original working title for this collection was *A User's Guide to Heavy Tails*, a title which was only dropped when we felt that there was some danger that it would primarily appeal to kangaroo hunters. Nevertheless, we impressed on our contributors to always keep the elusive "user" in mind, and as a result we believe that the papers in this volume will go a long way in helping a practitioner who

encounters heavy-tailed data. They provide tools, examples of different approaches, and a lead into the applied literature.

In this spirit, the volume opens with a section on applications. The two main applications considered are in the areas of computer networking and financial and insurance modelling. Crovella, Taqqu and Bestavros present convincing evidence of the heavy-tailed nature of the size distributions of files sent over the World Wide Web, and discuss the implications of this for network traffic, a topic that is continued in a paper by Willinger, Paxson and Taqqu which discusses related structural modelling problems.

On the economic side, Müller, Dacorogna and Pictet discuss the importance of heavy tails in the analysis of high frequency financial data, and look at the problem of tail decay parameter estimation in this setting, while Mittnik, Rachev and Paoletta discuss some general questions of heavy-tailed modelling in financial markets. The problem of risk management, in insurance and other financial settings, is treated in a paper by Bassi, Embrechts and Kafetzani via the use of quantile information.

The second grouping of papers centers around the problem of time series analysis for heavy-tailed data. Adler, Feldman and Gallagher give a comprehensive introduction to "Box-Jenkins" modelling in the stable setting, including a large number of simulations to indicate what does, and what does not, work. Calder and Davis describe parameter estimation in the stable time setting, followed by Taqqu and Teverovsky who treat the important problem of estimating long range dependence in finite and infinite variance series. These papers are followed with a thought provoking article by Resnick, which discusses a number of unexpected surprises and problems related to non-linearities and heavy-tailed modelling.

One of the interesting aspects of working with heavy-tailed, infinite variance time series is that many of the techniques used on finite variance series carry through with amazing success, although the technical details (such as the asymptotic sampling distributions of parameter estimates) may change dramatically. This is a recurring theme in all of the papers in this section, and is taken up again by Mikosch, who looks at the behaviour of "periodogram" estimates from heavy-tailed data.

The section closes with an illuminating article on sampling based Bayesian inference for heavy-tailed time series by Ravishanker and Qiou.

The third section of the volume contains two papers on general parameter estimation problems in the heavy-tailed setting. Pictet, Dacorogna and Müller describe an analysis of tail index estimation through Monte-Carlo simulation of synthetic data, in order to evaluate several tail estimators available in the literature. Ultimately, they recommend a bootstrapped and jackknifed version of the well known Hill estimator. A different ap-

proach to tail index estimation is taken by Kogan and Williams, who recommend working with the empirical characteristic function, and who suggest a method of getting around the heavy computational problems usually associated with this approach.

Sections 4–6 focus on specific statistical and modelling problems in which heavy-tailed distributions or processes play a central rôle. McCulloch considers the general regression problem when the error distribution is stable, while LePage, Podgórski and Ryznar discuss two resampling techniques for multiple linear regression with heavy-tailed errors. One is based on resampling permutations of residuals to the least squares estimates while the second exploits random flip signs. Both techniques are used to develop effective statistical inference for regression in the heavy-tailed setting.

Two more focused papers on signal processing then follow. The first, by Tsakalides and Nikias, discusses the “direction of arrival” estimation problem – a classical signal/noise problem – in a setting of stable noise. Their approach is via maximum likelihood estimation, which restricts their model to the Cauchy case, when likelihoods can be explicitly computed via analytic formula. (More on this below.) The second paper in this area, by Tshirintzis, presents and analyses a model for heavy-tailed interference arising from multiple users in communications networks.

Three general types of models are then presented by Goldie and Klüppelberg, Rosiński, and Samorodnitsky, who treat, respectively, subexponential distributions, the structure of stationary Lévy-stable processes, and shot noise processes with heavy-tailed shocks. These three papers, taken together, provide a solid insight to the structure of stable processes, and give a good indication of the wealth of models that exist in this area.

The volume closes with four papers related to the numerical aspects of stable distributions, two each by McCulloch and Nolan. There is no question that the development of fast and accurate numerical methods for computing stable densities is one of the main issues facing heavy-tailed modeling today.

Since the introduction of stable models, the impracticability of computing stable densities has been one of the main reasons for the need to develop non-standard statistical techniques in this setting. One could not, for example, employ the all but ubiquitous maximum likelihood techniques of standard (i.e. Gaussian) statistical analysis when there was no practical way of computing a likelihood. Today, with the advent of ever faster computers and new numerical techniques this possibility is close to being realised, and the fact that we have no analytic form for the stable density may soon no longer be a problem.

In his two papers in this closing section, McCulloch discusses the gen-

eral problem of numerical approximation of the symmetric stable distribution and density, and presents some tables for the maximally skewed case. Nolan discusses approximation, estimation, simulation and identification problems for multivariate stable distributions, and, in a short but important paper on numerical methods, gives us a URL for a battery of useful computer programs.

While reiterating once more the applied nature of this volume, it is important to note that many of the questions posed in the individual papers will require heavy theoretical analysis to be fully answered. Consequently, although we did not plan it this way, we rather expect that it will make a good source book for theoreticians as well, re-emphasising once again that the best theory is usually born in an application.

Finally, as editors, we have two sets of acknowledgements to make. The first is to our authors and referees. They all worked very hard to prepare papers that were useful and readable, rather than just "clever", as we are all trained to do nowadays. We take this opportunity also to apologise to them for all the rewriting we demanded.

Secondly, we must thank our granting agencies: RA is indebted to the Israel Science Foundation, the US-Israel Binational Science Foundation, the Office of Naval Research and, most recently, the National Science Foundation for support. RF thanks the Office of Naval Research. MT thanks the National Science Foundation.

Robert J. Adler
Raisa E. Feldman
Murad S. Taqqu

January, 1997.